README for ChatGPT and Public Sector Occupations Datasets (V.20240530)

Jieshu Wang (jwang490@asu.edu), Arizona State University

May 30, 2024

Table of Contents

1. Overview *1*

2. Datasets Included *2*

3. Dataset Descriptions *2*

3.1. 01_Public_sector_occupational_core_tasks_info_estimated_by_GPT_202405.csv **2**

3.2. 02_Public_sector_occupations_aggregated_info_202405.csv **3**

4. Usage Notes *4*

4.1. Data Source **4**

4.2. Potential Applications **5**

4.3. Contact Information **5**

1. Overview

This repository contains two datasets used in the study (listed below) exploring the impact of Generative AI, specifically ChatGPT, on the public sector workforce in the United States. The datasets provide detailed information on the core tasks of public sector occupations and their estimated performance metrics, including potential for automation and augmentation by ChatGPT. These estimations are generated by OpenAI's GPT-4 model (GPT-4-1106-preview) through OpenAI API.

- Wang, J., Kiran, E., S.R. Aurora (also known as Mai P. Trinh, Simeone, M., & Lobo, J. (2024). ChatGPT on ChatGPT: An Exploratory Analysis of its Performance in the Public Sector Workforce. *Digital Government: Research and Practice*.

2. Datasets Included

1. 01_Public_sector_occupational_core_tasks_info_estimated_by_GPT_202405.csv
2. 02_Public_sector_occupations_aggregated_info_202405.csv

3. Dataset Descriptions

3.1. 01_Public_sector_occupational_core_tasks_info_estimated_by_GPT_202405.csv

This dataset contains detailed information on 1,022 core tasks across 51 public sector occupations, including GPT-4's estimations of task performance metrics.

Columns:

General Information

- index: A unique identifier for each row in the dataset.
- O*NET-SOC Code: The code assigned by the O*NET system that corresponds to the Standard Occupational Classification (SOC) system. This code uniquely identifies the occupation within the ONET database.
- OCC_CODE: The occupation code used by the Bureau of Labor Statistics (BLS), which may slightly differ from the O*NET-SOC code.
- OCC_TITLE: The official title of the occupation.
- OCC_major_group_title: The major group title under which the occupation falls, according to the SOC classification.
- TOT_EMP: The total number of employees in this occupation, as reported by the BLS in May 2022.
- Job title: The specific job title for the tasks listed.

Task Information

- Task ID: A unique identifier for each task within the occupation.
- Task: A detailed description of the task performed in the occupation.
- task_importance: The importance rating of the task on a scale from 1 to 5, as derived from O*NET surveys.
- task_weight: The calculated weight of the task, based on its importance rating, used in aggregating task scores to the occupation level.
- DWA ID: The Detailed Work Activity ID, which categorizes specific activities performed within the task.

- IWA ID: The Intermediate Work Activity ID, which groups together related Detailed Work Activities.
- abilities_list: A list of abilities required for the task.
- skills_list: A list of skills required for the task.

GPT-4 Estimations (gpt-4-1106-preview)

- GPT_replacing_scale_1-10: The scale (from 1 to 10) estimating how much ChatGPT can replace human performance in this task, where 1 means no replacement and 10 means full replacement. Estimated by GPT-4.
- GPT_replacing_explanation: An explanation provided by GPT-4 justifying the replacement scale assigned.
- GPT_assistance_scale_1-10: The scale (from 1 to 10) estimating how much ChatGPT can assist humans in performing this task, where 1 means no assistance and 10 means full assistance. Estimated by GPT-4.
- productivity_increase: The estimated factor by which productivity can be increased with ChatGPT's assistance. For example, a value of 2 means the task can be completed in half the time with ChatGPT. Estimated by GPT-4.
- GPT_assistance_explanation: An explanation provided by GPT-4 justifying the assistance scale assigned.
- GPT_limitation: Details on the limitations of ChatGPT in performing or assisting with this task.
- human_relevance: An explanation of why human involvement remains crucial for this task.
- skills_to_maintain: Skills that humans need to retain to remain relevant in this task.
- skills_to_acquire: New skills that humans should develop to stay competitive in this task, given the presence of AI.
- obsolete_skills: Skills that are likely to become obsolete due to AI's capabilities in this task.

3.2. 02_Public_sector_occupations_aggregated_info_202405.csv

This dataset provides aggregated information on the 51 public sector occupations, summarizing the estimated performance metrics across all tasks within each occupation.

Columns:

General Information

- OCC_CODE: The occupation code used by the Bureau of Labor Statistics (BLS), which uniquely identifies the occupation.
- OCC_TITLE: The official title of the occupation.
- OCC_major_group_title: The major group title under which the occupation falls, according to the Standard Occupational Classification (SOC) system.
- TOT_EMP: The total number of employees in this occupation, as reported by the BLS.

Task Information

- core_task_count: The total number of core tasks identified for the occupation.

Ownership and Employment

- gov_ownership_pct_in_occupation: The percentage of employees within the occupation who are employed by federal, state, or local government entities.

Aggregated GPT-4 Estimations

- GPT_replacing_scale_1-10: The average scale (from 1 to 10) estimating how much ChatGPT can replace human performance across all core tasks in the occupation, where 1 means no replacement and 10 means full replacement. Aggregated based on task_weight.
- GPT_assistance_scale_1-10: The average scale (from 1 to 10) estimating how much ChatGPT can assist humans in performing tasks within the occupation, where 1 means no assistance and 10 means full assistance. Aggregated based on task_weight.
- productivity_increase: The aggregated productivity increase factor for the occupation, representing how much faster tasks can be completed with ChatGPT's assistance. For example, a value of 2 means tasks can be completed in half the time with ChatGPT's help. Aggregated based on task_weight.

4. Usage Notes

4.1. Data Source

The tasks and occupations are sourced from the O*NET database, which provides comprehensive information on job requirements and worker attributes in the United States. The estimations of GPT-4 are based on the GPT-4-1106-preview model.

4.2. Potential Applications

These datasets can be used for further research on AI's impact on workforce dynamics, policy-making, and developing strategies for integrating AI technologies into public sector operations.

4.3. Contact Information

For questions or further information, please contact:

- Name: Jieshu Wang
- Email: jwang490@asu.edu
- Institution: Decision Theater, Arizona State University.

We hope this data contributes to your research and understanding of AI's role in the public sector and future of work. Thank you for using our dataset!